

WHAT IS CLAIMED IS:

1. A network server load detection method comprising:  
a step of monitoring a communication from a client to a  
server, and counting a communication data size per connection  
5 as a load of said server;

a step of detecting a change in the communication data  
size per connection, and recording a maximum size value thereof;  
and

a step of judging, if the communication data size per  
10 connection decreases at this point of time with respect to the  
maximum size value, that said server is under a high load.

2. A network server load detection method according to  
claim 1, further comprising a step of counting a connection  
15 count and the communication data size till a count of  
communications monitored reaches a monitored communication  
minimum count and till a count time reaches a monitor minimum  
time by use of the monitored communication minimum count and  
the monitor minimum time.

20

3. A network server load detection method according to  
claim 1, further comprising a step of recognizing the  
communications of a start and end of the connection, and  
excluding communication data sizes of the start and end of the  
25 connection from a load detection target.

4. A network server load detection method according to

claim 1, further comprising:

a step of retaining information of the communication of the start of connection till the connection is ended or established;

5 a step of detecting the communication of the start of connection for re-connection executed when judging that said client fails to connect on the basis of the information retained; and

10 a step of setting a rate at which the communication of the re-connection occupies the number of the communications of the start of connection as a load of said server and, if this rate is high, judging that said server is under the high load.

5. A network server load detection method according to  
15 claim 1, further comprising:

a step of obtaining a distribution of the communication data sizes from said clients;

a step of distinguishing between extremely small pieces of communication data unrelated to the load of said server from  
20 the communication data size distribution; and

a step of eliminating the extremely small pieces of communication data from the judgement about the load.

6. A network server load detection method according to  
25 claim 1, further comprising:

a step of obtaining a sequence number from the communication to said server from said client;

a step of retaining a maximum value of the sequence number till the connection is ended since the start of connection;

a step of comparing the sequence number of the communication received with the sequence number retained: and

5 a step of excluding, if the sequence number obtained from the communication is smaller than the sequence number retained, this communication from counting.

10 7. A network server load detection method according to claim 1, further comprising:

a step of counting, if the sequence number obtained from the communication is smaller than the sequence number retained, the communication data after executing a weighting process thereon, or predicting a communication data size when there is  
15 no problem on a route from the two sequence numbers, and counting the predicted data size for detecting the load.

8. A network server load detection method comprising:

a step of monitoring a communication to a client from a  
20 server, and counting a receivable data size and a connection count of which said server notifies said client;

a step of obtaining the receivable data size per connection as a server load;

a step of storing a maximum value of the receivable data  
25 size per connection, and judging, if the receivable data size per connection becomes small with respect to the maximum value, that said server is under a high load.

9. A network server load detection system for monitoring a communication to a server from a client and detecting a load state of said server, comprising:

5        data size calculating means for calculating a size of communication data per connection;

         storage means for detecting a change in the communication data size per connection, and storing a maximum value; and

         load detection means for detecting a high load of said  
10        server when the communication data size per connection at that point of time with respect to the maximum value is equal to or smaller than a fixed value.

10. A network server sharing system for transferring data  
15        to a plurality of servers from a client via a network, comprising:

         routing means for transferring the data transmitted from said client to any one of said servers in a way of changing a destination of the data;

20        connection management means for retaining a mapping between the data and said server and indicating the destination to said routing means; and

         server sharing means for obtaining throughputs of said server, said client and a route by counting them, determining  
25        a correspondence between the data and said server by use of a function according to a service distribution rate based on the throughput, and transferring this correspondence to said

connection management means.

11. A network server sharing system according to claim 10, wherein said server sharing means sets, as the distribution rate, a modified probability distribution obtained by modifying a probability distribution corresponding to the throughput of said server so that the probability distribution is made more approximate to a uniform distribution as the throughputs of said client and of the route become lower.

12. A network server sharing system according to claim 10, wherein said server sharing means obtains a distribution of the throughputs of said client and of the route with respect to said client that is now on the service, also obtains a modified probability distribution by executing such a modification as to make the probability distribution corresponding to the throughput of said server more approximate to the uniform distribution as the throughputs of a new-connected client and of the route become lower for the distribution and reversely make the throughput of said server more outstanding as the throughputs of a new-connected client and of the route become higher, and sets this modified probability distribution as a distribution rate.

13. A network server sharing system according to claim 10, wherein a plurality of server sharing means are provided and each selected per client and service.